Energy and Buildings 80 (2014) 72–80

Contents lists available at ScienceDirect

Energy and Buildings

journal homepage: www.elsevier.com/locate/enbuild

Mid-term forecasting of urban electricity load to isolate air-conditioning impact

Luiz Friedrich, Peter Armstrong, Afshin Afshari*

Masdar Institute, Abu Dhabi, United Arab Emirates

ARTICLE INFO

Article history:
Received 12 December 2013
Received in revised form 10 April 2014
Accepted 2 May 2014
Available online 15 May 2014

Keywords:
Load forecasting
Regression
Energy efficiency
Retrofit
Measurement & verification
Air-conditioning load
Decision support

ABSTRACT

Demand Side Management (DSM) is often one of the most cost-effective approaches toward energy conservation and efficient electricity infrastructure utilization. Identifying total air-conditioning load for assessing targeted interventions is a difficult task given the transient thermal response of buildings, the coupled interaction of multiple sub-systems and the high correlation of demand with weather and other perturbations. An hourly regression-model of electricity consumption was developed for the city of Abu Dhabi, UAE, using measured hourly substation-level data. The fit is exceptionally good, even in a prediction context: Root Mean Squared Error (RMSE) equivalent to 1.54% of the annual peak load and Mean Absolute Percentage Error (MAPE) of 2.01% for the training period (full-year 2010), RMSE of 1.84% and MAPE of 2.64% for the testing period (first-half of 2011). The regression-model was combined with information from Abu Dhabi Urban Planning Council and the National Central Cooling Company (Tabreed) to derive an accurate estimate of the urban cooling load, the main driver of electricity consumption in the region. It was determined that, although only 30% of the annual load is directly weather dependent, air-conditioning explains no less than 57% of the total annual electricity load and 75% of the peak summer demand.

1. Introduction

According to the 2008 World Wildlife Fund's Living Planet Report [1], the UAE had the world's worst ecological footprint per person, experiencing only minor improvements by the publication of the 2012 report [2]. Demand-side energy efficiency is undoubtedly one of the most cost-effective ways for achieving Green House Gas emissions reduction.

The first step in promoting energy efficiency, from a policy point of view, should be to incentivize those initiatives that have the potential of achieving the highest impact in terms of energy cost savings with the lowest investment, the so-called "low hanging fruits". Identifying such options and estimating savings is a notoriously difficult task given the complexity and dynamics of the systems involved, the uncertain role of energy prices, lack of information on driving variables, unpredictability of end-user behavior and weather variability. Abu Dhabi and most of the region has a high portion of its electricity demand dedicated to cooling, mainly due to the hot and humid weather as well as the low thermal efficiency of the current building stock. The objective of this study was to identify, using different methods and sources of data, the fraction of the electricity load that can be directly traced back to cooling, since a series of measures are available to curb this specific driver of load.

2. Problem definition

Given the generally unsatisfactory level of thermal insulation in existing buildings and the extreme weather conditions during the long summer season, electricity load in the UAE and in most neighboring countries is highly correlated with weather for most of the year. It is estimated that the weather dependent portion of the load within Abu Dhabi municipality reaches 30% of the total load for the year, and 49% for the year's peak load hour [3]. The weather dependent load provides an absolute lower bound for the air-conditioning load, since the latter also includes certain constituents that are not weather dependent (e.g., pumps, fans). This fact reveals air-conditioning load as the primary target of any systematic energy efficiency program. To better model electricity consumption and the impact of air conditioning, it is essential to distinguish sensible cooling from latent cooling. Sensible cooling load refers to the energy demand directly responsible for keeping the building's indoor dry bulb temperature within a prescribed comfort range

* Corresponding author at: P.O. Box 54224, Abu Dhabi, United Arab Emirates. Tel.: +971 2 810 9149.
E-mail address: aafshari@masdar.ac.ae (A. Afshari).

http://dx.doi.org/10.1016/j.enbuild.2014.05.011
0378-7788/

Nomenclature

E	electricity load
Gl	linear growth
Gm	multiplicative growth
T	temperature
t	time
T_S	smoothed temperature
κ	forgetting factor
I	solar irradiance DNI_H
ω	solar irradiance coefficient
θ	composite temperature
α	ratio for composite temperature
$\widehat{Y}$	modeled hourly energy consumption
$\widehat{Y}_1$	modeled non-weather dependent fraction of load
$\widehat{Y}_2$	modeled weather dependent fraction of load
B	baseload
λ	coefficients for non-weather dependent parameters
F^a	Fourier series for annual seasonality
F^d	Fourier series for daily seasonality
D^f	dummy variable for Friday/Holiday
D^s	dummy variable for Saturday
D^r	dummy variable for Ramadan
θ	sensible cooling load proxy
γ	latent cooling load proxy
w	wind
$\widehat{\eta}\theta$	coefficient for sensible cooling
$\widehat{\eta}_\gamma$	coefficient for latent cooling
$\widehat{\eta}_w$	coefficient for wind
C	total load from chillers
P	Total load from pumps
F	total load from fans
$misc$	other sources of load
pC	percentage weather dependent chillers load
pP	percentage weather dependent pumps load
pF	percentage weather dependent fans load
$\bar{W}$	weather dependent load from regression model
$\widehat{W}$	weather dependent load from typical building
$\bar{B}$	non-weather dependent load from regression model
$\widehat{B}$	non-weather dependent load from typical building

(e.g., $24 \pm 2\,^{\circ}C$), while the latent cooling load is the energy required to keep the indoor air humidity within a prescribed comfort range (e.g., $55 \pm 5\%$). This is achieved through dehumidification of both the fresh air intake and the recirculating indoor air. The ultimate goal is to ensure that the indoor comfort conditions, usually defined in terms of temperature and relative humidity, are "acceptable" with reference to a certain standard (e.g., ASHRAE Standard 55 in North America).

The main objective of this study was to identify the energy consumed for indoor climate control, defined by heating, ventilation and air conditioning (HVAC) load. In order to identify such estimate a top-down mathematical model of the Abu Dhabi electricity consumption was developed and combined with information from a bottom-up model, which gives further information about the cooling load. The measured data provided by Abu-Dhabi's utility was separated in a way that two-thirds of it was used for the model training and one-third for model testing. We ensured that the low-complexity linear model (some light non-linearity explored) estimated coefficients are statistically significant and accurately represent the underlying physical phenomena.

In our model, temperature was used together with humidity, wind speed and solar irradiance to account for different weather influences. Regression analysis was the method chosen for combining and pondering the different effects, due to its relatively low computational cost and broad range of application. This approach also produces easily interpretable sensitivity coefficients for the different drivers of the load, facilitating the analysis of each component individually, as opposed to other top-down approaches where the physical significance of the coefficients is lost. The regression model was used to segment load into weather dependent and non-weather dependent component. In order to have a more detailed breakdown of electricity consumption, especially for identifying the base cooling load that is constant through the year (chiller base load, pumps and fans), the top-down model was combined with a bottom-up model representing the three main building types in Abu-Dhabi.

3. Literature study

There are two main classes of methods for predicting and analyzing aggregate urban energy usage [4], the bottom-up and the top-down approach.

A review of both classes of methods and comparison based on data requirements, pros and cons is presented in [5]. The traditional bottom-up approach to model urban energy use is usually based on data obtained from surveys and field measurements of the energy consumption profile and the characteristics of an individual unit or a statistically representative sample of units, identifying fixed demand per unit floor area or per household and extrapolating the results to infer global urban usage characteristics. This approach can give specific details on optimization potentials for a single building [6] but, by itself, usually fails to model the total energy use in buildings at urban level with sufficient resolution, due, among other shortcomings, to a non-linear relationship between load and floor area.

A more accurate approach, which can be combined with the top-down approach, consists of a series of typical building models used together in a way to represent the overall building stock. Individual simulation of each building focusing on energy use characteristics, are aggregated together based on the overall representativity of each building type in the studied area in order to predict urban energy use [7–9]. This approach can produce better results as compared to the traditional method, but some weaknesses arise. The results from the aggregate urban model are highly dependent on the precision of the individual building prototype models used, in the accuracy of the collected data on the representativity of each building in the urban building population, and in the assumptions made with regard to changing demographic factors, hours of occupancy, indoor climate control system in use etc. [10]. The work described in [11] shows the application of an urban energy model based on individual representative buildings combined into "building clusters" for specific districts and extrapolated to determine potential CO_2 emission reduction for Osaka city, Japan, considering different energy efficiency and energy saving measures. Because disaggregate data is usually obtained by surveys, precision and size of the underlying building database directly impact the precision of bottom-up models [12].

Due to the interaction between buildings and the heat stored in the urban canyon, an urban microclimatic phenomena known as the Urban Heat Island effect cannot be properly accounted for and estimated only by using isolated building models, since in practice the overall urban energy consumption is different from the sum of all its constituent building loads [13,14]. Given the complex dynamics of the system, the non-linearity and coupling of sub-systems and the high correlation with weather and other random perturbations, the bottom up approach has many limitations when applied to urban/district energy models. On the other hand, it does not

explicitly require the measurement of the overall urban energy consumption data—although such data, if available, may be used to calibrate the model. In addition, the bottom up model may be able to provide consumption details that are not distinguishable from a high-level model.

When high level measured data, such as the aggregate energy supply, macroeconomic data such as gross domestic product (GDP), population growth, buildings construction/demolition rate and meteorological data is available, and an overall representation of the building stock is desired, the top-down approach produces better results and enables high level consumption characteristics and trends to be identified. The parameters of the resulting models are either sensitivity coefficients or lumped physical properties. For the top-down approach, contributions can be broadly classified in three main categories [15]: time series models, artificial neural networks (ANN), regression-based approaches.

Time series models are best suited for capturing the time dependence of the load, seasonal effects (summer, winter), day-of-the-week (working day, non-working day), calendar holidays, different hours of the day and trends in general [16]. Some region-specific studies using this method are noteworthy: energy consumption and economic growth for south-east-Asia [17]; energy production and consumption in northern Spain [18]; comparison of hourly load profiles for Brazil [19] and short-term day-ahead and week-ahead load forecasting for Spain [20].

Artificial Neural Networks (ANN) methods, often classified as "black-box" type models, automatically interpolate between the electricity load and hour, day-of-week, day-of-year and weather data in a training data set based on a predefined model structure. The presumed advantage is the ability to learn complex and non-linear relationships that are difficult to model with conventional techniques [21–24]. Under this representation, the physical phenomena being represented cannot be matched to the coefficients (weights) of the ANN, therefore the marginal impact of each input variable is not clear from the model coefficients. Furthermore, there is a risk of over-fitting the model to the training data, which can hinder the validity of the model when used outside of the training period.

Regression models utilize the strong correlation of load with relevant factors such as weather, hour-of-the-day and day-of-the-week. Work using this method has proceeded on two fronts, single-equation models and multiple-equation models with different equations for different hours of the day [25–28]. A linear regression model was used to estimate the elasticities of GDP, price and GDP per capita for the domestic and non-domestic Italy's energy consumption for long term forecasting and comparing it with available projections [29]. Other than the economic factors, many previous studies have looked at the correlation between temperature and electricity consumption [30–32].

4. Data description

4.1. Electricity data

The quality of the electricity model and the final estimates are dependent on the quality of the data used for calibrating and testing, as well as the assumptions embedded in the modeling process. This study relied on substation-level hourly electricity data measured by the SCADA system of Abu-Dhabi Emirate's electricity utility, as well as hourly weather data (including direct and global solar irradiance and wind speed) monitored by Masdar City's comprehensive weather station for the calendar year 2010 (January 1st–December 31st) and the first half of 2011 (January 1st–June 30th). Abu-Dhabi's electricity load is mostly imputable to buildings (residential, commercial and institutional). The contribution

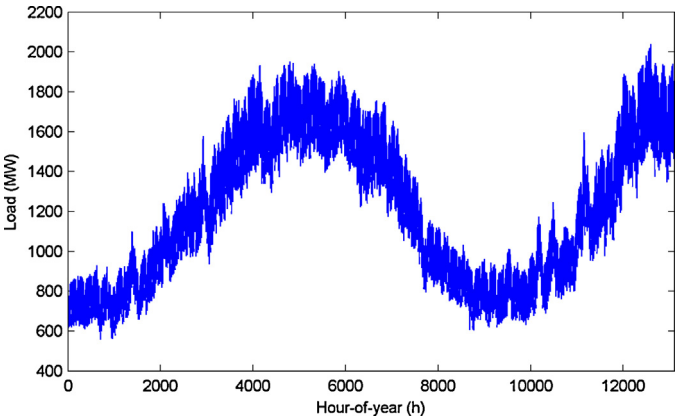

Fig. 1. Hourly load for 2010 and first half of 2011.

of the industrial & agricultural sectors is less than 12% of the total [33]. A subset of substations which are all within the municipality of Abu-Dhabi was selected, thereby eliminating industrial & agricultural loads. This subset constitutes, in aggregate, a better proxy for the buildings load than the total system load. Hourly electricity consumption data from 29 low-voltage substations (11 kV, 22 kV and 33 kV) was used. This subset includes all substations that directly supply street transformers serving final customers in residential/business areas within Abu-Dhabi municipality. The aggregate annual load of the selected substations peaks at 1935 MW in 2010, presenting a slight growth in 2011. Fig. 1 presents the plot of electricity demand (MW) per hour-of-year, starting from the 1st of January of 2010, which was used for model training and testing. To facilitate the interpretation of the model parameters, we normalize the load by dividing the hourly load measurements (in MW) by the peak demand observed in 2010.

4.2. Weather data

The Gauss–Markov theorem states that the least squares estimators have the smallest standard deviation of any unbiased estimator. This does not ensure that the standard deviation is always small in an absolute sense. In presence of multicollinearity [34], the least squares estimator maintains its properties but the standard deviation of the coefficients may be large, resulting in a non-efficient estimation. In order to minimize multicollinearity in the model, among the different solar irradiance measurements (Global Horizontal Irradiance – GHI, Diffuse Horizontal Irradiance – DHI and Direct Normal Irradiance – DNI) and derived DNI values (Horizontal DNI and Vertical DNI as presented in [35]), Horizontal DNI was determined to be the most significant of all irradiance measurements, and was the only variable representing solar irradiance retained in this study.

In addition to Horizontal DNI (DNI_H in W/m^2), dry-bulb temperature (Temp in °C), specific humidity (SH in g/g) and wind speed (Wind in m/s) showed significant correlation with electricity load (Load in MW). Fig. 2 presents the correlation between variables and the histogram distribution of each variable. Clearly, a certain degree of multicollinearity among the selected drivers cannot be avoided (Table 1).

4.3. Building models

To establish a business-as-usual benchmark of Abu Dhabi's energy and water consumption levels for application to the Pearl Design System Rating Method, Abu Dhabi Urban Planning Council (UPC) has established energy consumption benchmarks for the three most prevalent building types within Abu Dhabi Island,

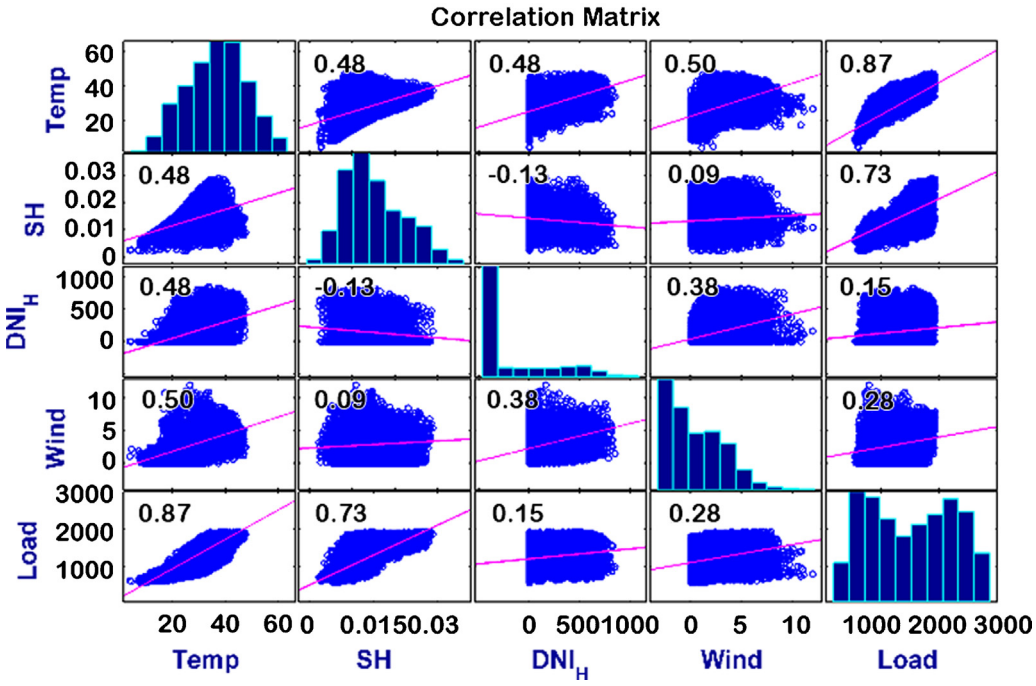

Fig. 2. Matrix plot, displaying variables distribution (charts on the diagonal) and cross-correlations.

(ARUP, private communication). In order to set such benchmark, different sources of information were used such as: end-user electricity meter information from Abu Dhabi Distribution Company (ADDC), building footprints and number of floors obtained from UPC's GIS department, walk-in and walk-by inspection and surveys, together with government reports from the Executive Affairs Authority (EAA) and the Regulation & Supervision Bureau (RSB). Three single-use representative building types were identified: Multi-unit residential, office and retail. Although villas represent a significant portion of the residential load in the emirate of Abu Dhabi, within the region covered by our substation selection, multi-unit residential buildings are prevalent; therefore, the UPC characterization remains valid. The UPC used the collected information to calibrate three detailed building models corresponding to the retained typologies. Finally, load constituents (chillers, pumps, fans, lighting, equipment and other services) for each building type were reported on a monthly basis. This information was further used to determine the contribution of air-conditioning to the aggregate Abu Dhabi load.

5. Prediction method

Our main contribution is to develop, via inverse modeling, a mid-term prediction model (able to predict up to a year ahead) of hourly electricity load for Abu Dhabi. Mid-term load prediction is to be distinguished from load forecasting, since the former is performed off-line (ex post) on historical time-series data (typically one year or more). It is the most suitable approach in order to establish a baseline model of electricity consumption. Furthermore, the model parameters represent the sensitivity of the load to the drivers of the load (exogenous variables) and can be used to predict the impact of planned large-scale DSM interventions on the aggregated urban electricity load.

Electricity load can be decomposed into two main constituents: a weather independent portion driven by calendar variables (annual seasonality, day-type and hour of the day) and a weather dependent portion where the main drivers are dry-bulb temperature, humidity, solar irradiance and wind speed.

The proposed model does not include ARIMA (Auto-Regressive Integrated Moving Average) components, as they are difficult to interpret from an economic/physical point of view and tend to result in uncontrolled growth of initial condition errors in a mid-term prediction setting. Fourier series were used to model daily (workday) and annual seasonalities, while dummy variables were used to differentiate Fridays/Holidays, Saturdays and Ramadan days from workdays (24 dummy variables per day type).

Abu Dhabi has seen significant population growth averaging 7.7% p.a. between 2005 and 2010 [33], leading to commensurate energy consumption growth in the last decade. To account for the non-stationarity of the data series resulting from this trend in electricity consumption through the year, two methods were tested: an additive linear trend and a multiplicative trend. The additive linear trend is based on the assumption that the rate of population inflow, new buildings and new appliances result in a constant hour-over-hour increase in the electricity consumption (Gl) throughout the year. This method considers that at every hour there is a constant load increment over the previous hour, independently of other factors. Defining $E(t)$ as the electricity load for hour t, $X(t)$ as the vector of exogenous variables affecting load at hour t and β the vector of coefficients for each variable, Eq. (1) represents the additive linear trend model. The second method using the multiplicative trend is based on the assumption that there is a rate of population inflow, new buildings and new appliances, which affects all parameters in the model (Gm). Eq. (2) presents the relation for the multiplicative

Table 1
Mean cross-correlation coefficients.

	Temp	SH	DNI_H	Wind	Load
Temp	1.00	0.48	0.48	0.50	0.87
SH	0.48	1.00	−0.13	0.09	0.73
DNI_H	0.48	−0.13	1.00	0.38	0.16
Wind	0.50	0.09	0.38	1.00	0.28
Load	0.87	0.73	0.16	0.28	1.00

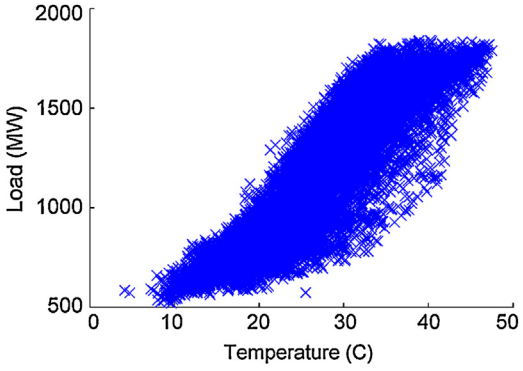

Fig. 3. Load versus temperature.

growth. Both growth methods were tested in order to determine the best representation of the physical behavior.

$$E(t) = Gl \times t + X(t) \times \beta \quad \forall t \tag{1}$$

$$E(t) = \left(1 + \frac{Gm \times t}{8760}\right) \times X(t) \times \beta \quad \forall t \tag{2}$$

6. Model definition

6.1. Aggregate urban model

The scatter plot of hourly load versus hourly temperature for 2010 (Fig. 3) clearly shows that there exists a non-linear relationship between temperature and load. A similar relationship was observed between load and enthalpy, the proxy identified to best represent ventilation load. This type of non-linearity has often been reported in the literature (e.g., [36]) especially in countries where the load is highly dependent on temperature, such as France (massive use of electric heating) or Spain (air-conditioning). In Abu-Dhabi, temperatures below a threshold of approximately 18 °C do not seem to have an impact on load while higher temperatures result in increased electricity demand [35]. Instead of using a piece-wise linear change-point model to account for the non-linearity between load and sensible/latent cooling proxies (e.g., [37]), a "threshold function" similar to the one implemented in the *Eventail* load model of Electricité de France, described by [15] was used, where a gradual transition around the change point is applied. This transformation applied to the sensible cooling proxy and to the latent cooling proxy account for the non-linear relation of temperature and enthalpy to load [3]. It also expresses, in a mathematically elegant way, the uncertainty associated with the threshold variable. The threshold function $\psi(x, x_0, \sigma)$ of a random variable x is the expected value of $\max(x - \chi, 0)$, where χ, the threshold variable, is normally distributed with mean x_0 and standard deviation σ. The transformation based on ψ corresponds to a left-truncated (one-sided) normal distribution function [34]. In order to account for the non-linear relation to load, and having eliminated the non-linearity, this transformed dataset is used in the linear regression model.

We now want to address another limitation of standard linear regression: its inability to represent dynamic (transient) effects. Mainly because of the thermal mass of buildings, it is known that [38] the cooling load shows a delayed response to weather variations—in particular temperature, but also solar radiation to some extent. To account for this thermal transfer inertia, the concept of a *smoothed* temperature is used as a proxy of the temperature felt inside the premises [15]. This method introduces a memory effect into the transformed smoothed temperature, whereby the latter becomes directly proportional to the instantaneous sensible cooling load, including a solar gain component ωI explained below. The smoothed temperature Ts is defined by Eq. (3) as an exponentially weighted moving average filter. The value of the forgetting factor κ has significant impact on the forecasting accuracy of the model. In our investigation, the same value suggested in [15] i.e., 0.98, equivalent to a filter time constant of $\tau = 50$ h ($\kappa = e^{-1/\tau}$), produced the best results for the Abu-Dhabi case.

$$T_S(t) = \kappa T_S(t-1) + (1-\kappa)(T(t) + \omega I) \tag{3}$$

Solar gain through external building surfaces often accounts for a major portion of the total envelope thermal gains [39]. The notion of sol–air temperature is an improvement over the simplified practice of accounting only for the difference between indoor temperature and outdoors ambient dry-bulb temperature [40,41], since it adjusts the outdoor temperature to account for solar radiation. The sol–air temperature can be defined as the equivalent outdoor temperature which will cause the same rate of heat flow at the surface in the absence of solar irradiation as the one that results from the aggregate effect of outdoor air temperature and the net radiation exchange between the surface and its environment [42], as represented in Eq. (4).

$$T_{\text{sol-air}} = T + \frac{\alpha I - \Delta Q_{ir}}{h_o} \tag{4}$$

where: T, outdoor temperature. α, average solar radiation absorptivity of the surfaces. I, solar irradiance incident on the surfaces. ΔQ_{ir}, extra infrared radiation due to difference between external air temperature and apparent sky temperature. h_o, Heat transfer coefficient for radiation and convection.

To account for the solar gain, a component with horizontal DNI (I) was included in the definition of the smoothed temperature, since we determined earlier that it presents the highest correlation with load. As for the optimal value of ω, the multiplier for the horizontal DNI component in the smoothed temperature, a value of 13 °C m^2/kW was determined through line search minimization of the prediction error. The convective heat transfer coefficient h_o depends on wind speed. In order to account for this influence, we add a separate *Wind* parameter to the regression model of the load, assuming linear relationship for simplicity.

The smoothed temperature represents the heat transfer inertia through building walls and roof, but there is also an instantaneous impact, for example through air infiltration or mechanical ventilation. Therefore, a composite temperature θ is defined by Eq. (5), which combines the smoothed temperature with the instantaneous temperature. Once again, the optimal value of α was determined, through line search minimization of the prediction error, to be 0.19, thereby giving considerable weight (more than 80%) to the smoothed temperature.

$$\theta = (1-\alpha)T_S + \alpha T \tag{5}$$

The approach using the composite temperature θ produced a cleaner final model, as a single coefficient is identified for the combined effect of all weather variables affecting the overall Abu-Dhabi's sensible cooling load. The accuracy of the model is also significantly enhanced by this approach.

The model is specifically formulated to deal with the triple seasonality that typically arises in electricity data. This triple seasonality includes daily, weekly and annual cycles. Daily and annual seasonalities are modeled using fourth order Fourier Series (8 parameters per seasonality), while for the weekly seasonality four day types are considered: Workday (same as daily seasonality), Saturday, Friday/Holiday and Ramadan. Each of the non-workday daily profiles is characterized by a set of 24 dummy variables. The load displays a monotonic trend—monotonically increasing in this case.

Let $\widehat{Y}$ represent the hourly energy consumption after growth was accounted for, the overall model formulation is shown in Eq.

(6), where $\widehat{Y}_1$ represents the non-weather dependent fraction of the load and $\widehat{Y}_2$ the weather dependent fraction, *for each hour*. In total, the calendar-driven portion of the load (Eq. (7)) is represented by 8 annual seasonality parameters + 8 daily seasonality parameters + 24 Friday/Holiday hourly dummy variable parameters + 24 Saturday hourly dummy variable parameters + 24 Ramadan hourly dummy variable parameters = 88 exogenous variables to which we must add a constant term B which we call "baseload", representing the yearly average non-weather dependent load. The non-weather dependent portion of the load at any given point in time is baseload + annual seasonality + Workday daily seasonality + Friday/Holiday offset (if this is a Friday or a Holiday) + Saturday offset (if this is a Saturday) + Ramadan offset (if this is a Ramadan day). Note that these offsets can be cumulative. For instance, on a Ramadan day which is also a Friday/Holiday or a Saturday, both relevant offsets need to be added to the standard Workday seasonality.

In Eq. (8), η is the vector of coefficients corresponding to the non-calendar exogenous drivers of the load, while X_2 is the vector of measured/derived exogenous variables. Eq. (9) shows the coefficients of the non-calendar exogenous variables, where θ represents the sensible cooling, γ the latent cooling (enthalpy) and w wind component, while Eq. (10) shows the constituents of the vector of exogenous variables, X_2.

$$\widehat{Y} = \widehat{Y}_1 + \widehat{Y}_2 \tag{6}$$

$$\widehat{Y}_1 = B + \lambda \times \left[F_1^a \cdots F_8^a F_1^d \cdots F_8^d D_1^f \cdots D_{24}^f D_1^s \cdots D_{24}^s D_1^r \cdots D_1^r\right]' \tag{7}$$

$$\widehat{Y}_2 = \eta \times X_2 \tag{8}$$

$$\eta = \left[\hat{\eta}_\theta \, \hat{\eta}_\gamma \, \hat{\eta}_w\right]' \tag{9}$$

$$X_2 = \left[\theta \, \gamma \, w\right]' \tag{10}$$

The additive growth term *Gl* expresses a linear relationship between load and time and can therefore be estimated simultaneously with the other parameters of the regression model. In order to identify *Gm*, the multiplicative trend, without resorting to non-linear estimation, the coefficients of the linear model were identified for all growth rates from 0% to 10% with 0.5% increments (alternatively, a nonlinear parameter identification procedure can be used to determine all model parameters simultaneously). The growth rate which produced the best fit for the overall model over the training period, was then used to test its prediction power over the testing period.

Electricity price elasticity is usually accounted for in similar studies, since price fluctuations often produce significant changes in electricity consumption [29]. Abu Dhabi has seen stable electricity prices in the past years; therefore, the impact of price elasticity cannot be inferred from the data available. Furthermore, for reasons beyond the scope of the present study, the pricing level is unlikely to be used in a DSM context within the emirate of Abu Dhabi in the foreseeable future. Therefore, price elasticity was not considered in the present study, but could be an important addition to future work as electricity price variation happens.

The model estimation procedure resulted in auto-correlated residuals. This is a phenomenon that is often encountered in load forecasting studies. Due to the nature of the problem, and the objectives of the study, this problem could not be directly solved, nor is it a major impediment to the accuracy of our mid-term predictions. It merely indicates that our linear regression model, despite the adjustments described above, does not yet completely account for all sources of non-linearity and non-stationarity. Under the Gauss–Markov assumption the residuals are assumed to be independently distributed and non-correlated to each other. The Ordinary Least Squares (OLS) framework is known to under-estimate the parameter estimation uncertainties when this assumption is violated, affecting the validity of the *t*-statistics. In order to overcome this problem, an identification method robust to autocorrelation and heteroskedasticity in the error terms was used. The Newey–West estimator [43] provides an estimate of the covariance matrix of the parameters to improve the OLS estimation of the errors.

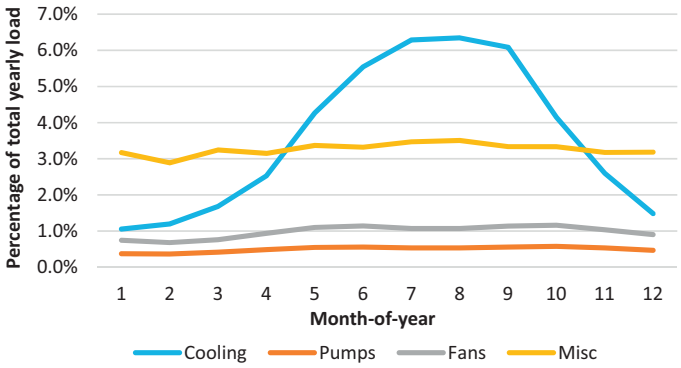

Fig. 4. Cooling, pumps, fans and *misc* load profiles in % of year load.

Different models were tested with a combination of the variables and methods previously discussed, in order to identify the best representation of the overall electricity load behavior. The models were compared over the training period using the Root Mean Squared Error (RMSE) and the Mean Absolute Percentage Error (MAPE). The best models were further tested to measure their accuracy over the testing period. The RMSE and the MAPE were used to assess the model's prediction capability against the measured data from the first half of 2011. A summary of this comparison is presented in Section 7.

6.2. Building typology model

The regression model allowed load to be separated into weather dependent and non-weather dependent components. However, as mentioned above, the weather dependent load is actually a lower bound for the cooling load. In order to have an accurate estimate of the cooling load, an analysis using UPC typical building profiles was conducted. For each building type, UPC has determined a typical electricity consumption profile, breaking it down in Cooling (representing the chillers load), pumps, fans and *misc*.

According to estimates from Statistics Center Abu Dhabi [33], 45% of the non-industrial load is due to residential customers while the remainder can be attributed to commercial, office and institutional buildings. This information was used to produce an average building profile based on the three representative building types from UPC: 45% weight for the residential type, and 27.5% each for office and retail types. The average electricity consumption profile is presented in Fig. 4. Considering that for each building type, the total electricity load due to cooling (*C*), pumps (*P*) and fans (*F*) is composed of a weather dependent portion as well as a portion that is not dependent on weather, the weather driven portions of cooling, pumps and fans loads are designated as *pC*, *pP* and *pF* respectively. The remaining portions $(1 - pC)$, $(1 - pP)$ and $(1 - pF)$ are not dependent on weather. Fig. 5 represents this relation; the label "*misc*" identifies all load components that are not directly related to cooling.

The weather dependent percentages (*pC*, *pP*, *pF*) were identified by comparing the weather dependent load obtained from the regression model $\bar{W}$ to the estimated weather dependent load from the typical buildings in Abu Dhabi $\widehat{W}$, presented in Eq. (11). Similarly, the non-weather dependent load from the regression model $\bar{B}$ was compared to the portion of the typical buildings' load that is

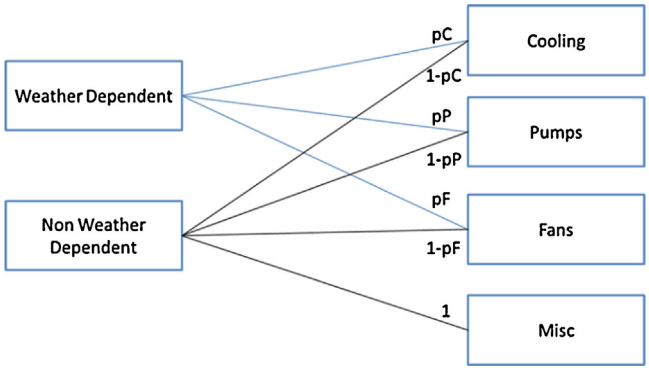

Fig. 5. Weather and non-weather dependent loads.

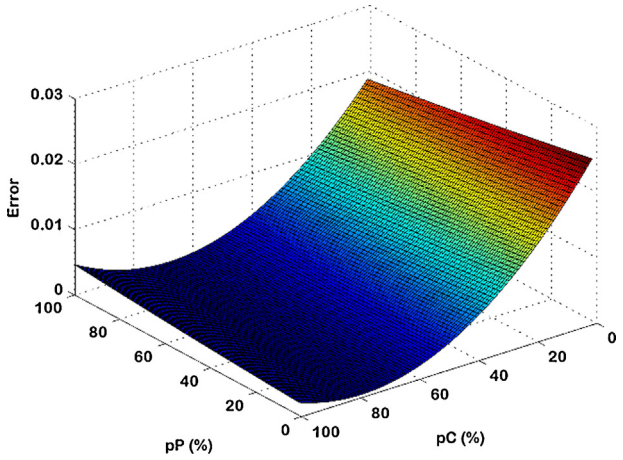

Fig. 6. Fitting of *pP* and *pC* to the modeled data.

not related to cooling. Eq. (12) presents this relation. Using a line search minimization routine, the summation of the squares of the differences $\bar{W} - \widehat{W}$ and $\bar{B} - \widehat{B}$ was minimized (Eq. (13)), in order to identify the percentage of the non-weather dependent component of the load from the regression model that is actually attributable to the energy use of certain cooling equipment. For instance, Fig. 6 shows the surface of the error e based on pP and pC. The optimal point identified indicates the 79% of the chiller energy is weather dependent, while 100% of pumps and fans load is independent from weather and fully accounted for in the annual seasonality component of the regression model.

Table 2
Summary of training period results.

Growth method	Optimal growth (%)	RMSE	MAPE
Additive	0.94	1.544	1.99
Multiplicative	3	1.541	2.01

Table 3
Weather dependent coefficients.

	Coefficient	Std. error	*t*-Stat
Sensible	0.958	0.035	26.9
Latent	0.203	0.01	19.9
Wind	−0.429	0.038	−11.3

$$\widehat{W} = pC \times C + pP \times P + pF \times F \tag{11}$$

$$\widehat{B} = (1 - pC) \times C + (1 - pP) \times P + (1 - pF) \times F + misc \tag{12}$$

$$\text{minimize } e = \sum (\widehat{W} - \bar{W})^2 + \sum (\widehat{B} - \bar{B})^2 \tag{13}$$

7. Results and discussion

For the training period, both growth methods produced similar results, very high adjusted *R*-squared (0.993), and RMSE and MAPE as presented in Table 2. However, the multiplicative growth model generally behaves better in a prediction setting (testing period) and was therefore retained for this analysis. In the testing period, it is often necessary to adjust the growth rate; this is feasible given that, in our target application, said period actually lies in the past and the corresponding growth rate can be inferred from publicly available statistics in a straightforward manner. For the testing period, the multiplicative growth term was adjusted to 11% giving the best results with RMSE of 1.841. Despite some autocorrelation of the residuals, the model was able to capture the seasonality and growth trend in the data as presented in the residual plot and histogram distribution for the model for the training period in Figs. 7 and 8.

The quasi-totality of the estimated parameters was statistically significant (as per the *t*-statistics provided by the Newey–West algorithm). Table 3 presents the coefficients and standard error for the weather dependent parameters. The coefficient of the proxy for sensible cooling load is about 5 times that of the proxy for latent cooling load. Wind speed, used as a parameter in the model

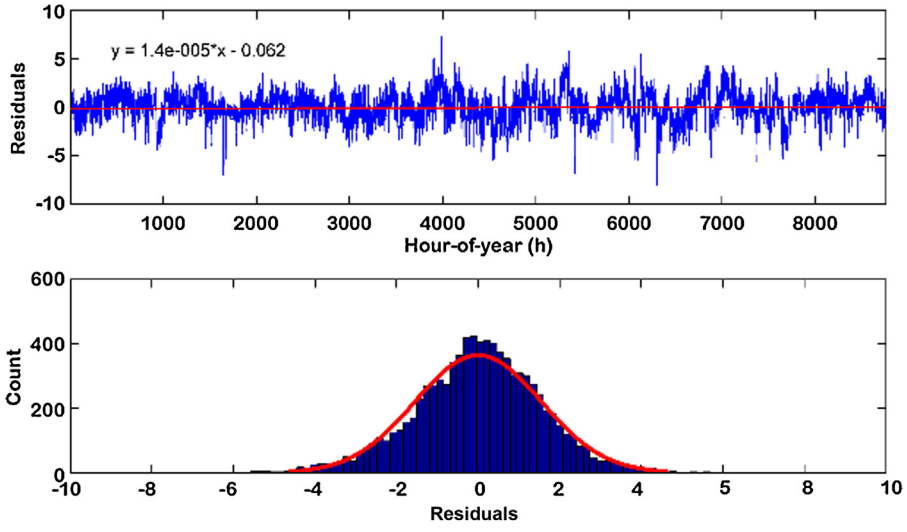

Fig. 7. Model training residuals (a) and histogram (b).

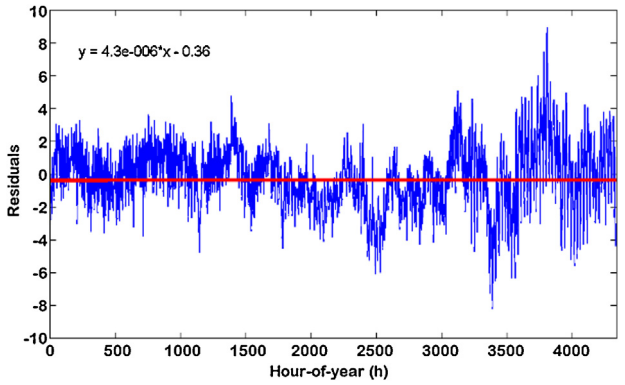

Fig. 8. Model testing residuals.

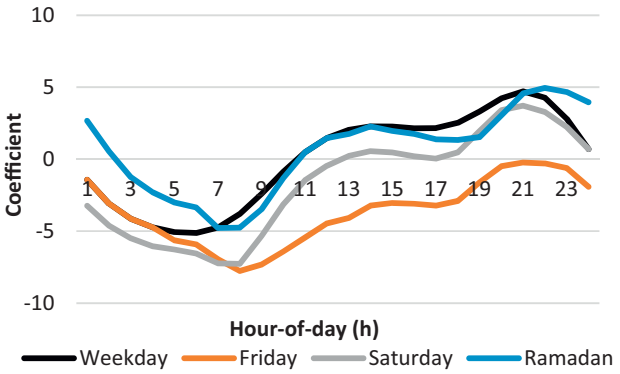

Fig. 9. Daily seasonality profile per day-type.

to account for the convection rate in the sol–air temperature, is surprisingly significant (cooling effect), although the impact is not comparable with temperature or humidity. A few dummy variable coefficients mainly pertaining to certain hours of the day for Ramadan day-types that were not significant were removed (i.e., set to zero). Fig. 9 presents the daily profile for each day type. It can be observed that on Friday the base load is lower through the day, especially in the afternoon period as most businesses are closed. Saturday exhibits a smaller overall reduction on baseload as only some offices are closed on that day. During Ramadan, a large portion of the population is fasting during daytime, and have gatherings in the evening for the ceremonial breaking of the fast. Dummy variables representing Ramadan represent this behavior, where baseload during nighttime is substantially higher compared to other days. Fig. 10 presents the annual seasonality profile.

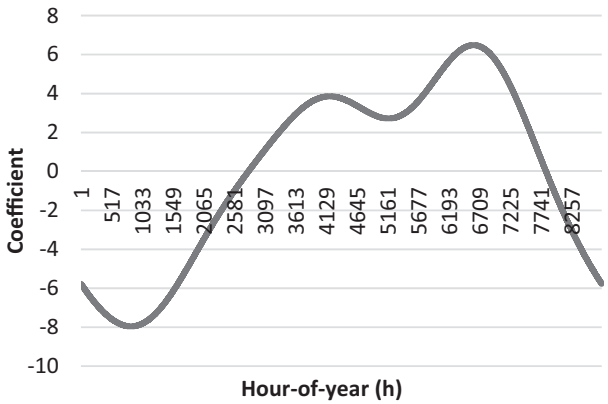

Fig. 10. Annual seasonality profile.

8. Conclusions

Identifying the potential of air-conditioning related Demand Side Management measures and building retrofits is a difficult task given the nonlinear and dynamic nature of the thermal response of buildings, the coupled interaction of multiple sub-systems and the high correlation of cooling demand with weather and other perturbations. The lack of detailed measured data on end energy usage specifically of air-conditioning and its drivers increases the uncertainty for policy makers targeting energy efficiency. Measured substation level hourly data was used to develop and identify a regression-model of electricity consumption for the city of Abu Dhabi, UAE, and this model combined with Urban Planning Council data to determine cooling demand for the Abu-Dhabi Municipality.

The model can be used for DSM program design optimization via a priori scenario generation. Specifically, the sensitivity of the annual load and peak demand to the variation of some of the regression coefficients representing building physical features, can give a reasonably accurate indication of the expected city-wide impact of large-scale DSM interventions aimed at modifying said features. UPC typical building profiles were combined to the regression model here described to identify the total cooling load, as some of the cooling components are accounted for in the non-weather dependent fraction of the load. It was identified that 79% of the chiller energy is weather dependent, an extra 21% of the non-weather dependent load is used for chillers and should be accounted for in the air-conditioning load. Similarly, the load used for pumps and fans are entirely included in the non-weather dependent load, under the annual seasonality, and should therefore be considered as part of the air-conditioning load. The National Central Cooling Company (Tabreed) has reported that on average the cooling plants in the region operate at a 20% baseload compared to the summer peak demand (private communication). The base load during winter can be considered to constitute the non-weather dependent cooling requirement, which is very close to the 21% non-weather dependent chiller load that we identified. By combining the 30% weather dependent load with the baseload chiller's consumption, as well as the pumps and fans when estimating the global air-conditioning load, the annual load reaches 57% of the total electricity consumption. For the peak hour of the year, the contribution of the air-conditioning load reaches 75%, compared to the 49% when only weather-dependent load is considered. The application of this knowledge may assist urban planning and policy making, resulting in lower operational costs and GHGs emissions, as well as development of future building codes for new and existing buildings.

Future work will focus on simultaneous estimation of the regression linear coefficients, the coefficients defining the composite temperature and the multiplicative growth term. This will increase the accuracy of the model, at the expense of an increase in computational complexity. Furthermore, a more dynamic modeling approach (transfer function, state-space) might be able to fully utilize the information present in the data, while producing physically significant lumped parameters of the average building—conductive/convective/solar gain, chiller efficiency, air tightness, etc.—which can be used for advanced sensitivity analysis and more accurate and specific what-if scenario analysis. Finally, the explicit consideration of the Urban Heat Island effect, as well as the long-wave radiation losses will result in a better characterization of the urban energy consumption. In the more distant future, electricity price variations and elasticity will be investigated in order to gain better insights for scenario analysis and policymaking.

Acknowledgements

The authors acknowledge the assistance of the Abu Dhabi Water and Electricity Authority, for the provision of sub-station level hourly load data and the Abu Dhabi Urban Planning Council for the detailed specification report describing the Emirate's building typologies.

References

[1] WWF, Living Planet Report 2008, 2008.
[2] WWF, Living Planet Report 2012, 2012.
[3] L. Friedrich, An Urban Energy Baseline Model for Measurement & Verification of Building Energy Efficiency Retrofits in Abu-Dhabi (Master's thesis), Masdar Institute of Science and Technology, Abu Dhabi, UAE, 2013.
[4] L. Suganthi, A.A. Samuel, Energy models for demand forecasting—a review, Renewable and Sustainable Energy Reviews 16 (2) (2012) 1223–1240.
[5] L.G. Swan, V.I. Ugursal, Modeling of end-use energy consumption in the residential sector: a review of modeling techniques, Renewable and Sustainable Energy Reviews 13 (8) (2009) 1819–1835.
[6] M. Palonen, A. Hasan, K. Siren, A genetic algorithm for optimization of building envelope and HVAC system parameters, in: Proc. IBPSA'09, 2009.
[7] J. Huang, H. Akbari, L. Rainer, R. Ritschard, 481 Prototypical Commercial Buildings for 20 Urban Market Areas, US Department of Commerce, National Technical Information Service, 1991.
[8] P. Jones, S. Lannon, J. Williams, Modelling building energy use at urban scale, in: 7th International IBSPA conference, Rio de Janeiro, Brazil, August, 2001, pp. 1683–1697.
[9] J. Clarke, et al., Simulation support for the formulation of domestic sector upgrading strategies, in: Proceedings of 8th International IBPSA conference, 2003, pp. 219–226.
[10] M. Kavgic, A. Mavrogianni, D. Mumovic, A. Summerfield, Z. Stevanovic, M. Djurovic-Petrovic, A review of bottom-up building stock models for energy consumption in the residential sector, Building and Environment 45 (7) (2010) 1683–1697.
[11] Y. Yamaguchi, Y. Shimoda, M. Mizuno, Proposal of a modeling approach considering urban form for evaluation of city level energy management, Energy and Buildings 39 (5) (2007) 580–592.
[12] L. Shorrock, J. Dunster, The physically-based model BREHOMES and its use in deriving scenarios for the energy use and carbon dioxide emissions of the UK housing stock, Energy Policy 25 (12) (1997) 1027–1037.
[13] M. Kolokotroni, I. Giannitsaris, R. Watkins, The effect of the London urban heat island on building summer cooling demand and night ventilation strategies, Solar Energy 80 (4) (2006) 383–392.
[14] B. Bueno, L. Norford, G. Pigeon, R. Britter, A resistance-capacitance network model for the analysis of the interactions between the energy performance of buildings and the urban climate, Building and Environment 54 (2012) 116–125.
[15] A. Bruhns, G. Deurveilher, J.-S. Roy, A non linear regression model for mid-term load forecasting and improvements in seasonality, in: Proceedings of the 15th Power Systems Computation Conference, 2005, pp. 22–26.
[16] N. Amjady, Short-term hourly load forecasting using time-series modeling with peak load estimation capability, IEEE Transactions on Power Systems 16 (3) (2001) 498–505.
[17] J. Asafu-Adjaye, The relationship between energy consumption, energy prices and economic growth: time series evidence from Asian developing countries, Energy Economics 22 (6) (2000) 615–625.
[18] S.G. Chavez, J.X. Bernat, H.L. Coalla, Forecasting of energy production and consumption in Asturias (northern Spain), Energy 24 (3) (1999) 183–198.
[19] L.J. Soares, M.C. Medeiros, Modeling and forecasting short-term electricity load: a comparison of methods with an application to Brazilian data, International Journal of Forecasting 24 (4) (2008) 630–644.
[20] J.R. Cancelo, A. Espasa, R. Grafe, Forecasting the electricity load from one day to one week ahead for the Spanish system operator, International Journal of Forecasting 24 (4) (2008) 588–602.
[21] D.C. Park, M. El-Sharkawi, R. Marks, L. Atlas, M. Damborg, and others, "electric load forecasting using an artificial neural network", IEEE Transactions on Power Systems 6 (2) (1991) 442–449.
[22] C.-N. Lu, H.-T. Wu, S. Vemuri, Neural network based short term load forecasting, IEEE Transactions on Power Systems 8 (1) (1993) 336–342.
[23] T. Chow, C. Leung, Neural network based short-term load forecasting using weather compensation, IEEE Transactions on Power Systems 11 (4) (1996) 1736–1742.
[24] P. Mandal, T. Senjyu, T. Funabashi, Neural networks approach to forecast several hour ahead electricity prices and loads in deregulated market, Energy Conversion and Management 47 (15) (2006) 2128–2142.
[25] R. Cottet, M. Smith, Bayesian modeling and forecasting of intraday electricity load, Journal of the American Statistical Association 98 (464) (2003) 839–849.
[26] V. Dordonnat, S.J. Koopman, M. Ooms, Dynamic factors in periodic time-varying regressions with an application to hourly electricity load modelling, Computational Statistics & Data Analysis 56 (11) (2012) 3134–3152.
[27] R. Ramanathan, R. Engle, C.W.J. Granger, F. Vahid-Araghi, C. Brace, Short-run forecasts of electricity loads and peaks, International Journal of Forecasting 13 (2) (1997) 161–174.
[28] T. Reddy, J. Kissock, S. Katipamula, D. Claridge, An energy delivery efficiency index to evaluate simultaneous heating and cooling effects in large commercial buildings, Journal of Solar Energy Engineering (United States) 116 (2) (1994) 79–87.
[29] V. Bianco, O. Manca, S. Nardini, Electricity consumption forecasting in Italy using linear regression models, Energy 34 (9) (2009) 1413–1421.
[30] P.J. Robinson, Modeling utility load and temperature relationships for use with long-lead forecasts, Journal of Applied Meteorology 36 (5) (1997) 591.
[31] F. Apadula, A. Bassini, A. Elli, S. Scapin, Relationships between meteorological variables and monthly electricity demand, Applied Energy 98 (2012) 346–356.
[32] A. Pardo, V. Meneu, E. Valor, Temperature and seasonality influences on Spanish electricity load, Energy Economics 24 (1) (2002) 55–70.
[33] SCAD, Statistical Yearbook of Abu Dhabi 2011, 2011.
[34] W.H. Greene, Econometric Analysis, 2003.
[35] M.T. Ali, M. Mokhtar, M. Chiesa, P. Armstrong, A cooling change-point model of community-aggregate electrical load, Energy and Buildings 43 (1) (2011) 28–37.
[36] M. Bessec, J. Fouquau, The non-linear link between electricity consumption and temperature in Europe: a threshold panel approach, Energy Economics 30 (5) (2008) 2705–2721.
[37] J. Kelly Kissock, C. Eger, Measuring industrial energy savings, Applied Energy 85 (5) (2008) 347–361.
[38] P.R. Armstrong, S.B. Leeb, L.K. Norford, Control with building mass—Part I: Thermal response model, ASHRAE Transactions 112 (1) (2006).
[39] T.H. Kuehn, J.W. Ramsey, Thermal Environmental Engineering, 1998, pp. 474–481.
[40] D.G. Erbs, S.A. Klein, W.A. Beckman, Sol–air heating and cooling degree-days, Solar Energy 33 (6) (1984) 605–612.
[41] P.W. O'Callaghan, S.D. Probert, Sol–air temperature, Applied Energy 3 (4) (1977) 307–311.
[42] K.R. Rao, Some Investigations on the Sol–Air Temperature Concept, Commonwealth Scientific and Industrial Research Organization, Australia, Melbourne, 1970.
[43] W.K. Newey, K.D. West, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, Econometrica: Journal of the Econometric Society 55 (3) (1987) 703–708.